

CLAIMS

- 1 1. A method for comparing a first content with a second content to determine
2 whether the contents are identical, the method comprising the steps of:
3 identifying a protocol encoding the first content and second content;
4 computing a first signature of the first content and a second signature of the sec-
5 ond content; and
6 comparing the first computed signature with the second signature to determine
7 whether the first content is identical to the second content.
- 1 2. The method of claim 1 further comprising the steps of:
2 selecting a first set of data segments from the first content and a second set of data
3 segments from the second content; and
4 using the selected first set of data segments and the second set of data segments to
5 compute the first signature and the second signature.
- 1 3. The method of claim 2 wherein the selected first set of data segments and second
2 set of data segments comprise locations associated with one or more protocol markers.
- 1 4. The method of claim 1 wherein the step of computing the signature of the first
2 content and the signature of the second content further comprises the steps of:
3 identifying one or more protocol markers associated with the first content; and
4 identifying one or more protocol markers associated with the second content.
- 1 5. The method of claim 4 wherein the one or more protocol markers associated with
2 the first content comprises discrete cosine coefficients.
- 1 6. The method of claim 4 wherein the one or more protocol markers associated with
2 the second content comprises discrete cosine coefficients.

- 1 7. The method of claim 4 wherein the one or more protocol markers associated with
2 the first content comprises motion vectors.
- 1 8. The method of claim 4 wherein the one or more protocol markers associated with
2 the second content comprises motion vectors.
- 1 9. The method of claim 4 further comprising the steps of:
2 identifying a length of the first content; and
3 identifying a length of the second content.
- 1 10. A content comparator adapted to compare first content with a second content, the
2 comparator comprising:
3 a protocol identification module configured to identify a first protocol associated
4 with the first content and a second protocol associated with the second content;
5 a plurality of data segmentation modules configured to select a set of data seg-
6 ments from each of the first content and the second content;
7 a plurality of signature computation modules configured to generate a first signa-
8 ture of the first content and a second signature of the second content; and
9 a signature comparison module configured to compare the first signature with the
10 second signature.
- 1 11. An apparatus for comparing a first content with a second content, the apparatus
2 comprising:
3 means for identifying a protocol encoding the first content and the second content;
4 means for selecting a set of data segments from the first content and the second
5 content;
6 means for computing a signature of the first content and a signature of the second
7 content; and
8 means for comparing the computed signature of the first content with the com-
9 puted signature of the second content.

1 12. The apparatus of claim 11 wherein the selected data segments comprises locations
2 associated with one or more protocol markers.

1 13. The apparatus of claim 11 wherein the means for computing the signature of the
2 first content and the signature of the second content further comprises:
3 means for identifying one or more protocol markers associated with the first con-
4 tent; and
5 means for identifying one or more protocol markers associated with the second
6 content.

1 14. The apparatus of claim 13 wherein the one or more protocol markers associated
2 with the first content comprises discrete cosine coefficients.

1 15. The apparatus of claim 13 wherein the one or more protocol markers associated
2 with the second content comprises discrete cosine coefficients.

1 16. The apparatus of claim 13 wherein the one or more protocol markers associated
2 with the first content comprises motion vectors.

1 17. The apparatus of claim 13 wherein the one or more protocol markers associated
2 with the second content comprises motion vectors.

1 18. The apparatus of claim 13 further comprises:
2 means for identifying a length of the first content; and
3 means for identifying a length of the second content.

1 19. A method to compare a first content with a second content in a network storage
2 environment, the method comprising the steps of:
3 receiving the first content;
4 computing a signature of the first content;

5 comparing the computed signature of the first content with a signature of the second
6 content; and

7 identifying, if the computed signature of the first content matches the signature of the
8 second content, that the first content is identical to the second content.

1 20. The method of claim 19 wherein the step of computing the signature of the first
2 further comprises the steps of:

3 identifying a set of protocol markers associated with the content; and
4 generating the signature from the identified set of protocol markers.

1 21. The method of claim 19 wherein the set of protocol markers further comprise a set
2 of discrete cosine coefficients.

1 22. The method of claim 19 wherein the set of protocol markers further comprises
2 one or more motion vectors.

1 23. The method of claim 19 wherein a size of the received content is utilized in cre-
2 ating the signature.

1 24. A method for identifying content using a protocol associated with the content as a
2 signature, the method comprising the steps of:

3 determining the protocol associated with the content;
4 identifying a set of markers associated with the protocol;
5 obtaining a set of markers from the content using the set of marker associated
6 with the protocol; and
7 generating a signature of the content using the identified markers.

1 25. The method of claim 24 wherein the identified markers are within a subset of the
2 entire content.

- 1 26. The method of claim 24 wherein a size associated with the content is utilized to
2 uniquely identify the content.
- 1 27. The method of claim 24 wherein the signature is utilized in a network caching de-
2 vice to determine whether data should be forwarded from the network caching device.
- 1 28. The method of claim 24 wherein the signature is utilized to determine if a local
2 copy of the content should be accessed.
- 1 29. The method of claim 24 wherein the signature is utilized to determine if a remote
2 copy of the content should be accessed.
- 1 30. A protocol maker identifier for generating a signature of a content comprising:
2 a data segmentation module configured to select a set of data segments from the
3 content; and
4 a signature computation module configured to generate the signature from the set
5 of data segments.
- 1 31. The protocol maker identifier of claim 30 further comprising a protocol identifi-
2 cation module configured to identify a protocol associated with the content.
- 1 32. The protocol maker identifier of claim 30 wherein the signature comprises a set of
2 protocol markers.
- 1 33. The protocol marker identifier of claim 32 wherein the set of protocol markers
2 comprises a set of discrete cosine transform coefficients.
- 1 34. A network caching device adapted to utilize a signature associated with a protocol
2 for caching decisions, the network caching device comprising:
3 means for determining a protocol of new contents;

4 means for computing a signature of the content; and
5 means for comparing the computed signature of the new content with a signature
6 of other content.

1 35. The network caching device of claim 34 wherein the means for computing a sig-
2 nature further comprises:

3 means for identifying a set of makers associated with the protocol associated with
4 the content; and

5 means for obtaining appropriate markers associated with the content.